Early Online

The European Research Journal 2025 Original Article

DOI: https://doi.org/10.18621/eurj.1641267 *Medical Physics and Deep Learning*

Pneumonia detection in chest X-ray images using convolutional neural networks

Çağdaş Şimşek[1], Suat Özkorukcuklu[1], Bora Işıldak[2]

[1]Deparment of Physics, İstanbul University, Faculty of Science, İstanbul, Türkiye; [2]Deparment of Physics, Yıldız Technical University, Faculty of Art and Science, İstanbul Türkiye

ABSTRACT

Objectives: Pneumonia ranks among the infections and presents a considerable health threat, especially in certain age groups and developing countries. The accurate diagnosis of the disease and prompt identification are crucial for treatment purposes. This study aimed to develope a convolutional deep neural network model that can detect pneumonia using a sufficient number of chest X-ray images that have been verified with a "definite diagnosis" clinically.

Methods: This study uses a dataset that includes 1000 chest X-ray images from a variety of age groups taken as part of patient care at Koç University Faculty of Medicine Hospital Clinics. The dataset sample includes two sets of pictures called normal and pneumonia infected. Various preprocessing techniques were used on the obtained images, thus enabling the training and testing of our developed prediction model.

Results: We improved the accuracy of the model's decisions by applying image processing techniques, successfully achieving high levels of decision accuracy with our model We have elevated the precision of decision-making in our model to outstanding levels and achieved impressive F1 Score and AUC (Area Under the Curve) values (F1 Score: 0.94 and AUC Score: 0.98).

Conclusions: Our model was trained using X-ray images produced from the same devices of the same hospital and achieved very high prediction results, but using images produced from different countries, different hospitals and different devices, especially training and testing the model with much larger data sets, is a necessary need for this study and the model we developed to become more universal, and in this sense, there is a need to develop and expand the study.

Keywords: Pneumonia, convolutional neural network, chest X-ray, deep neural network, machine learning, deep learning, artificial intelligence

Studies have shown that pneumonia is a reason for hospital admissions in people of all ages and can be a serious health concern for the elderly population as it may result in severe illness and even death. A research investigation, into the number of deaths associated with pneumonia has revealed that national mortality statistics from 14 years of 19 european union countries were analysed and reported to be the most common cause of infection-related deaths, especially among the elderly and individuals with comorbidities. [1]. Children under 5 ages and individuals over 65 ages are, at risk of pneumonia related complications.

Received: February 18, 2025 **Accepted:** April 17, 2025 **Available Online:** July 23, 2025 **Published:** XX XX, 2025

How to cite this article: Şimşek Ç, Özkorucuklu S, Işıldak B. Pneumonia detection in chest X-ray images using convolutional neural networks. Eur Res J. 2025. doi: 10.18621/eurj.1641267

Corresponding author: Çağdaş Şimşek, PhD., Phone: +90 212 455 57 00, E-mail: cagdas.simsek@cern.ch

In 2019 2.5 million deaths globally were linked to pneumonia [2].

Lung infection known as pneumonia can be caused by microorganisms like bacteria or viruses and is identified by respiratory symptoms and lung abnormalities on imaging tests such as infiltrations or consolidations in lung tissue or pleural effusion [3, 4].

The diverse range of infections across societies and age groups highlights a need to understand potential causes for effective treatment strategies. Identifying pathogens that cause lower respiratory tract infections, especially in childhood, is a challenging process because disease indicators may overlap with respiratory pathogens not related to infection, and existing routine tests may often lead to false-negative results or detect randomly transmitted pathogens [5].
Commonly seen in viral pneumonia are similar symptoms that can be observed by the doctor during a physical checkup to evaluate the patient's general well-being including respiratory difficulties and mental alertness. The doctor also checks the signs such as temperature and pulse rate along, with examining lung abnormalities by inspecting the rib cage and listening to the patients' breathing sounds. The results provide clues that suggest pneumonia may be present but to confirm the diagnosis definitively a chest X ray is necessary [6].

Alveolar infiltration seen on a chest X ray is described as the buildup of cells and fluid within the alveoli (air sacs). This pattern is frequently seen in pneumonia caused by bacteria such, as Streptococcus pneumoniae. On an X ray image alveolar infiltration appears as a dense opacity.
A common feature of pneumonia is the presence of boundaries usually limited to a specific part of the lung (referred to as lobar pneumonia).

Another characteristic is infiltration which occurs when inflammation and fluid buildup happen in the connective tissue between the air sacs that support lung structure. Known as the interstitium. This type of infiltration is often seen in cases of pneumonia or pneumonia caused by less common microorganisms like Mycoplasma pneumoniae. X ray images show infiltration as a broader and more spread-out pattern, with fine reticular or nodular features; however, this distinction may not always be clear cut. Some bacterial pneumonias like Staphylococcus aureus pneumonia can show infiltration while certain viral pneumonias may display alveolar infiltration as well. It's important to consider the findings along, with clinical observations and lab tests to determine the cause of pneumonia correctly [7].

Getting a chest X ray is a radiographic procedure used for screening purposes. Diagnosing lung conditions is commonly done using chest X ray imaging due to its cost effectiveness and ease of access without invasive procedures involved in the process of identification from the images relies heavily on skilled and seasoned doctors' knowledge and training " This challenge becomes particularly significant in regions, with limited healthcare resources where both the frequency and severity of these disorders are higher.

Pediatric pneumonia rates are much higher than the global average figures suggest. Automated diag-

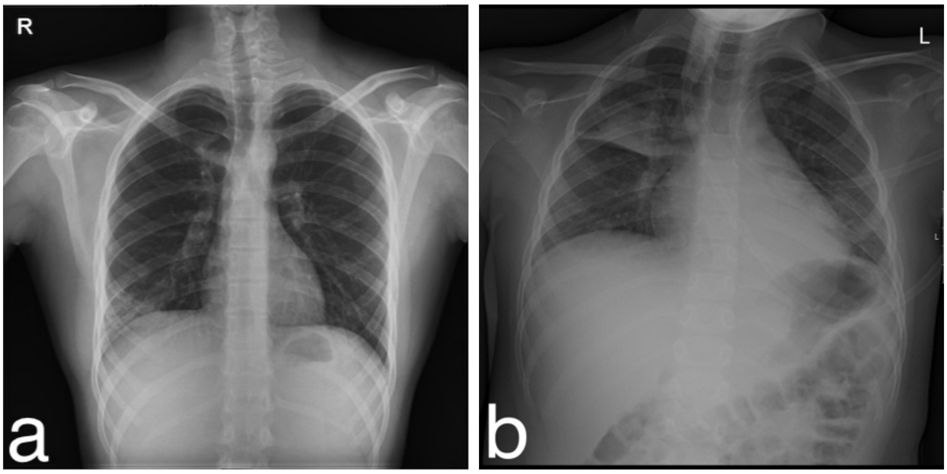

Fig. 1. **Chest X-ray of a) normal patient and b) pneumonia infected patient.**

nostic tools using chest X rays are seen as a way to improve radiologists' productivity and cut down on healthcare costs while speeding up the detection and treatment of pneumonia in kids.

METHODS

Experimental Dataset

This research study uses a dataset that includes 1000 chest X-ray images from a variety of age groups taken as part of patient care at Koç University Faculty of Medicine Hospital Clinics. The sample includes two sets of pictures called normal and pneumonia infected as shown in Fig. 1. Pneumonia is divided into bacterial and viral pneumonia. Qualified radiologists assigned labels during the data preparation process for patients, with a confirmed diagnosis based on chest X-ray images and advanced imaging methods as necessary. Backed up by laboratory tests and further validated through pathological examinations when needed and ultimately decided upon by the physicians.

The council's viewpoint was taken into consideration when organizing the imaging process and conducting laboratory procedures on samples from both healthy individuals and those with pneumonia infection for Convolutional Neural Network (CNN) training purposes later on; the images, in these sample subsets were then shuffled randomly for analysis as outlined in Table 1. Examining smaller portions, for analysis and evaluation.

Table 1. Splitting of the dataset

Category	Training sample	Test sample
Normal images	190	60
Pneumonia infected images	565	185
Total images	755	245
Percentage of total images	75%	25%

Our approach to identifying pneumonia is based on the predictions we make. In order to guarantee the reliability and accuracy of the predictions made by our CNN model we have incorporated measures, into the image preprocessing stage. One, after the other algorithms are executed in sequence.

Resize Pictures

The X-ray images of patients' chests can vary depending on factors like age and health status well as body type. In order to create a model, we meticulously adjusted all the images to ensure the lung, to image ratio is optimized. This manual adjustment led to images of sizes after cropping each one, as illustrated in Fig. 2. Therefore, we included the tactics mentioned below in our approach.

Preparing Images, for Analysis

After using the picture cropping method, we applied three processing steps to each image before including them in the training phase. Firstly, we adjusted all images to 512 × 512 size to match the CNN models

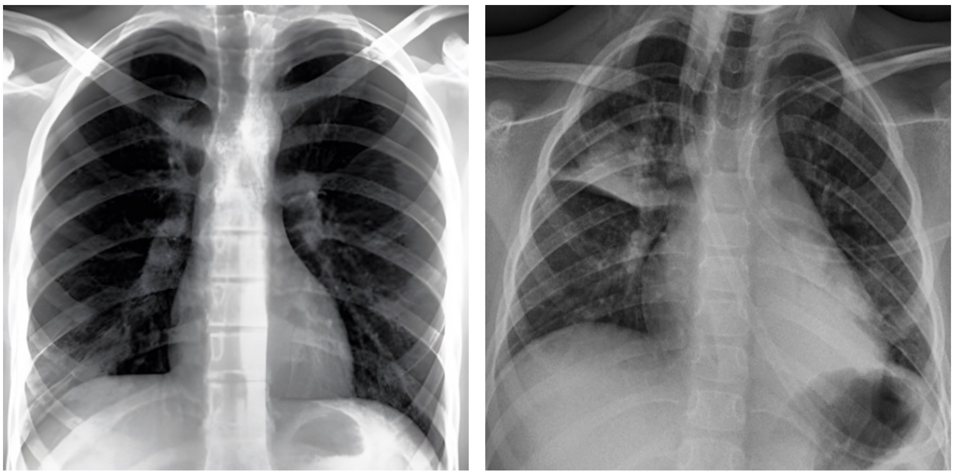

Fig. 2. Histogram equalized and cropped versions of the images shown in Fig. 1.

input requirements. We used bilinear interpolation to decrease the image sizes, for uniformity. In this step, we performed histogram equalization on the images to improve their overall quality and clarity.
Fig. 2 of the report is where you can see the effects of histogram equalization clearly shown in comparison to methods used in image processing tasks like normalizing pixel values within a specific range of 1 to 255 for better data scaling, in the third layer adjustment process.

Enhancing Data

We improved our sample by rotating it at an angle selected from a uniform range of [-10, 10] degrees. After the enhancement process was completed, we also doubled the size of the training sample set. Rotating images randomly boosts the model's robustness in accommodating the positioning of young patients and others. Severe illnesses can show images, on X-rays naturally.

Training at CNN

Many studies have shown results with various networks like ResNet18 [8], DenseNet121 [9], Xception [10], MobileNet [11]. However, these networks are designed for datasets. They have tens of millions of parameters for training A network with too many parameters trained with a small sample like ours often results in significant fluctuations, in loss values affecting accuracy values. Therefore the model cannot be considered.

Our study utilized a CNN design showcased in Fig. 3 for analysis purposes. During the training phase of the model development process running over 100 epochs with a batch size of 32. We opted for employing the square error (MSE) loss function paired with the Adam optimizer and maintained a learning rate set at 0.0001, during the training session.

Assessing Performance

The balanced datasets play a role in enhancing the predictive accuracy of both the Diagnostic and Diagnostic models in comparison to imbalanced datasets that require more complex neural network structures to effectively handle the data distribution disparities observed within the dataset where normal instances occur roughly four times less frequently than infected cases.

There seems to be a difference and inconsistency in the model predictions output observed here; however, if the model can maintain stability and accuracy even when trained under such conditions, at higher levels, it would give a significant edge in adjusting the model for various data sets broadly. The dataset comprises 93% of the information used for analysis purposes in this model's network configuration. When the number of layers and nodes increases, in the network model it tends to lead to a risk of overfitting, which can make the accuracy of the prediction algorithm more complex to achieve effectively.

Generating Grad-CAM Visualizations

In the realm of visual recognition tasks within

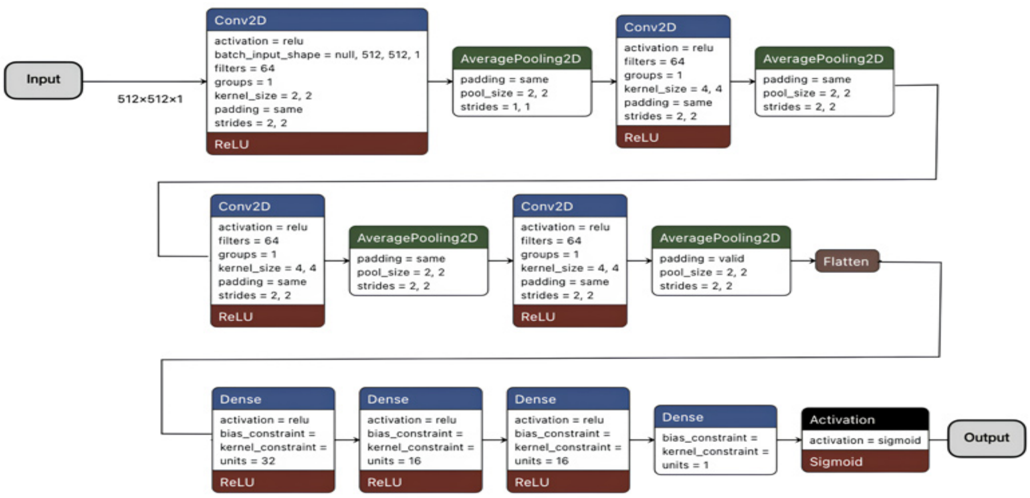

Fig. 3. Layer structure of Convolutional Neural networks with hyperparameters.

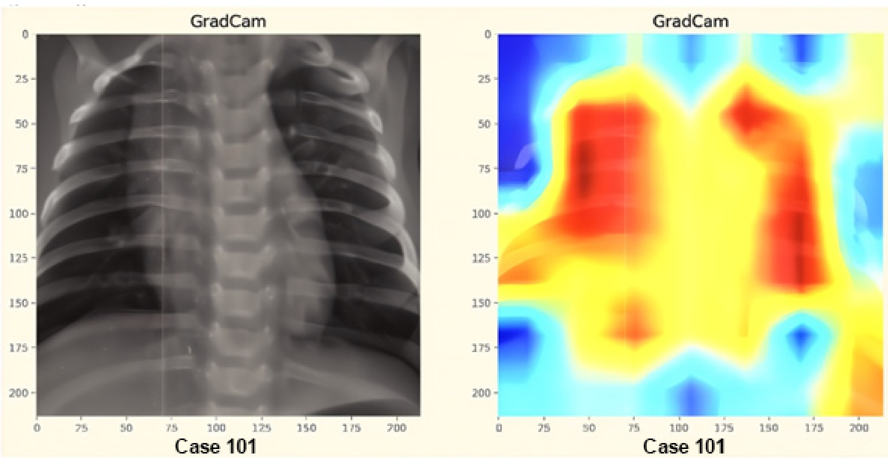

Fig. 4. When the model renders a judgment, the pixels it emphasizes are depicted in red.

deep learning models have achieved notable advancements over time. However understanding how these models make decisions proves to be a challenging endeavor. Shedding light on the characteristics that drive the model's decisions is crucial, for enhancing its reliability and pinpointing any shortcomings in this context.

Methods such as Grad CAM (Gradient weighted Class Activation Mapping) have emerged as a tool, for understanding how deep learning models work [12]. When applied to a model designed to detect pneumonia in chest X-rays, Grad CAM has demonstrated the model's ability to accurately identify pneumonia lesions [12]. We improved the clarity of the model's decisions for healthcare providers. Boosted its dependability, with Grad CAM technology that reveals which pixels the model focuses on most intensely in its decision-making process. This helps us better understand and improve the model's capabilities and weaknesses. We tested our model using a selected chest X-ray image and after it classified the image as either normal or abnormal.

The Grad CAM method was applied to analyze the image in Fig. 4 showcasing the areas where our model focused on while making decisions. The highlighted red regions indicate the key areas emphasized during the decision-making process.

Statistical Analysis

McNemar test was applied to evaluate the statistical significance of the classification performance of our model in pneumonia detection. As a result of the test using the number of false positives and false negatives in the test set, the test statistic was found to be 1.000 and the P-value was found to be 0.317.

This result shows that there is no statistically significant difference between the errors of the model (P>0.05). This shows that the performance of the model is balanced and not excessively biased towards a particular class. Therefore, the classification performance of the model is balanced and reliable.

RESULTS

In the sections for training and testing data sets the model accuracy results for both datasets are shown

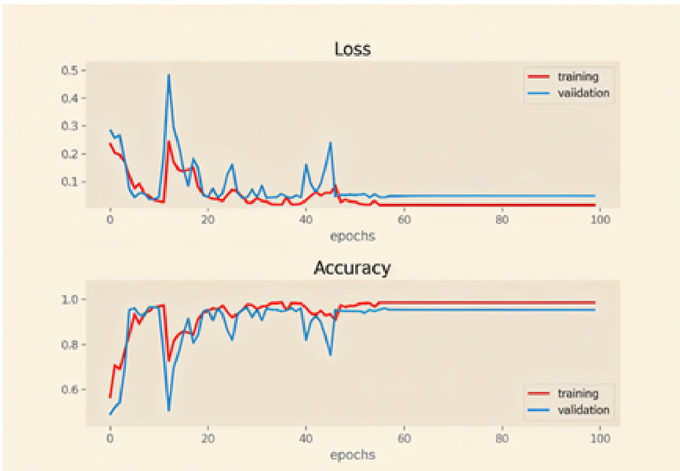

Fig. 5. Loss and accuracy curves for both training and validation sets during 100 epochs.

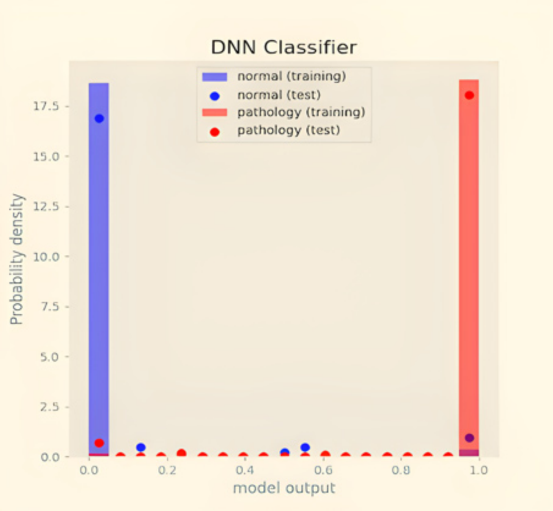

Fig. 6. CNN output distribution for normal images in the training sample (blue shaded), for the normal images in the test sample (blue dots), for pneumonia infected in the training sample (red shaded) and for pneumonia infected in the test sample (red dots).

below. The model attains 100% accuracy in the training set with its accuracy, in the test set being reached. The top part of Fig. 5 shows how the loss values change over training sessions. Fig. 6 shows the normalized output distributions of CNN for pneumonia-infected images, in both training and test sets are shown separately for each image type.

Table 2 displays the precision metrics for the model's performance assessment on the test sample using recall and F1 score values as well as the area under the Receiver Operating Characteristic (ROC) curve to quantify the obtained scores, in Fig. 7 left and right panels respectively.

DISCUSSION

Throughout the training phase of our model development, it's important to check for signs of overfitting to ensure performance. Taking a look at the distribution of model outputs for training and test data separately can help us assess whether the model is overfitting or not. Noticing a difference in the distributions of certain class labels can indicate potential issues with overfitting.

Inconsistencies between the training and test data could suggest that the model is overfitting the data for pneumonia detection. During the study period both pneumonia images from training and testing sets did not show any notable variances in results. The accuracy measure alone could have effectively assessed the performance of a model without overfitting had there existed a distribution of sample classes utilized in the training process. However, due to the absence of balance in our case, further metrics, like precision, recovery, F1 score, and AUC score have been incorporated for evaluation purposes.

The importance of the F-Score, especially in studies with dataset imbalances, is described in the study "A Survey of Evaluation Metrics for Classification Performance" [13]. The F-Score is important in areas such as medical diagnoses where the consequences of false negatives (for example, a cancer-positive case being described as cancer-negative) are crucial.

Limitations

In addition to all these, the limitations of the F1 score are discussed in one of the most important journals, the study titled "A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives", which is one of the current studies published in ACM Computing Surveys (CSUR). For example, it is mentioned that it ignores the performance of the negative class while evaluating the performance of the positive class, and this situation can be problematic especially in multi-class classification problems [14].

CONCLUSION

In order to build a reliable and quantitative CNN model, we used an appropriate amount of data from chest X-ray images with a high-test accuracy of (93.3%) and AUC score of (98.1%), using an adequate quantity of chest X-ray image data. The CNN model predictions

Table 2. Precision, recall, F1 score and AUC Score obtained for the test sample.

Pressicion	Recall	F1 Score	AUC Score
0.93	0.95	0.94	0.98

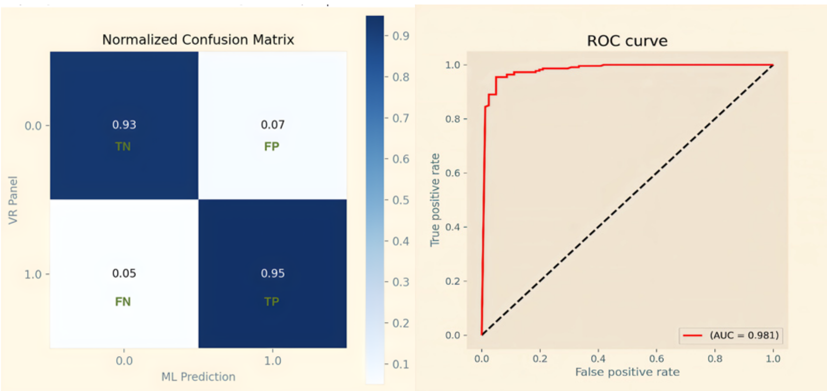

Fig. 7. **Confusion Matrix obtained for the test sample. TN, TP, FN, FP stand for True Negative, True Positive, False Negative, False Positive respectively (left). ROC curve of the CNN model. The area under this curve gives the AUC score (right).**

are meant to expand the information available to clinicians during the decision-making process, especially in the case of pediatric triage. The purpose of presenting a prediction generated by artificial intelligence is to provide additional information for the radiologist in the decision-making process. The diagnosis has to be made by a specialist. On the other hand, there are places in the world where there are not enough doctors who can correctly identify pneumonia from chest X-ray images. A simple application on a smartphone can be used to gather better diagnostic information in these regions using the CNN model we have developed.

However, our model was trained using X-ray images produced from the same devices of the same hospital and achieved very high prediction results, but using images produced from different countries, different hospitals and different devices, especially training and testing the model with much larger data sets, is a necessary need for this study and the model we developed to become more universal, and in this sense, there is a need to develop and expand the study.

Ethical Statement

This study was approved by the Koç University Biomedical Research Ethics Committee (Decision no. 2025.300.IRB2.141, date: 30.06.2025).

Authors' Contribution

Study Conception: SÖ, BI; ÇŞ; Study Design: SÖ, BI, ÇŞ; Supervision: SÖ; Funding: SÖ; Materials: ÇŞ; Data Collection and/or Processing: ÇŞ, BI; Statistical Analysis and/or Data Interpretation: ÇŞ, BI; Literature Review: ÇŞ; Manuscript Preparation: ÇŞ, BI and Critical Review: SÖ, BI.

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgement

All Chest X-ray images and the dataset used in this study were obtained from Koç University Faculty of Medicine Hospital.

I would also like to express my sincere gratitude to Prof. Dr. Mehmet Onur Demirkol for his valuable support and contributions in the clinical evaluation of medical images.

Generative Artificial Intelligence Statement

The author(s) declare that artificial intelligence tools were used in accordance with academic ethical standards during the preparation of this manuscript. Overleaf AI Assist tool was used in text editing operations such as checking spelling errors, editing words and adapting the reference source format. The all content of the study was produced by the author(s) in accor dance with scientific research methods and academic ethical principles.

Editor's note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Marshall DC, Goodson RJ, Xu Y, et al. Trends in mortality from pneumonia in the Europe union: a temporal analysis of the European detailed mortality database between 2001 and 2014. Respir Res. 2018;19(1):81. doi: 10.1186/s12931-018-0781-4.
2. World Health Organization (WHO). Pneumonia in children. 2022. cited 2025 Feb 20. Available from: https://www.who.int/news-room/fact-sheets/detail/pneumonia
3. Mandell LA, Wunderink RG, Anzueto A, et al; Infectious Diseases Society of America; American Thoracic Society. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44 Suppl 2(Suppl 2): S27-72. doi: 10.1086/511159.
4. Bartlett JG. Diagnostic tests for agents of community-acquired pneumonia. Clin Infect Dis. 2011;52 Suppl 4: S296-304. doi: 10.1093/cid/cir045.
5. Mick E, Tsitsiklis A, Kamm J, et al. Integrated host/microbe metagenomics enables accurate lower respiratory tract infection diagnosis in critically ill children. J Clin Invest. 2023 Apr 3;133(7):e165904. doi: 10.1172/JCI165904.
6. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med. 2019;200(7): e45-e67. doi: 10.1164/rccm.201908-1581ST.
7. File TM Jr, Ramirez JA. Community-Acquired Pneumonia. N Engl J Med. 2023;389(7):632-641. doi: 10.1056/NEJMcp2303286.
8. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi: 10.1109/CVPR.2016.90.
9. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected Convolutional Networks. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269. doi: 10.1109/CVPR.2017.243.
10. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800-1807. doi: 10.1109/CVPR.2017.195.
11. Howard AG, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017 Apr 17.
12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626. doi: 10.1109/ICCV.2017.74.
13. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process. 2015;5(2):1-11. doi: 10.5121/ijdkp.2015.5201.
14. Christen P, Hand DJ, Kirielle N. A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives. ACM Comput. Surv. 2024;56(3):73. doi: 10.1145/3606367.